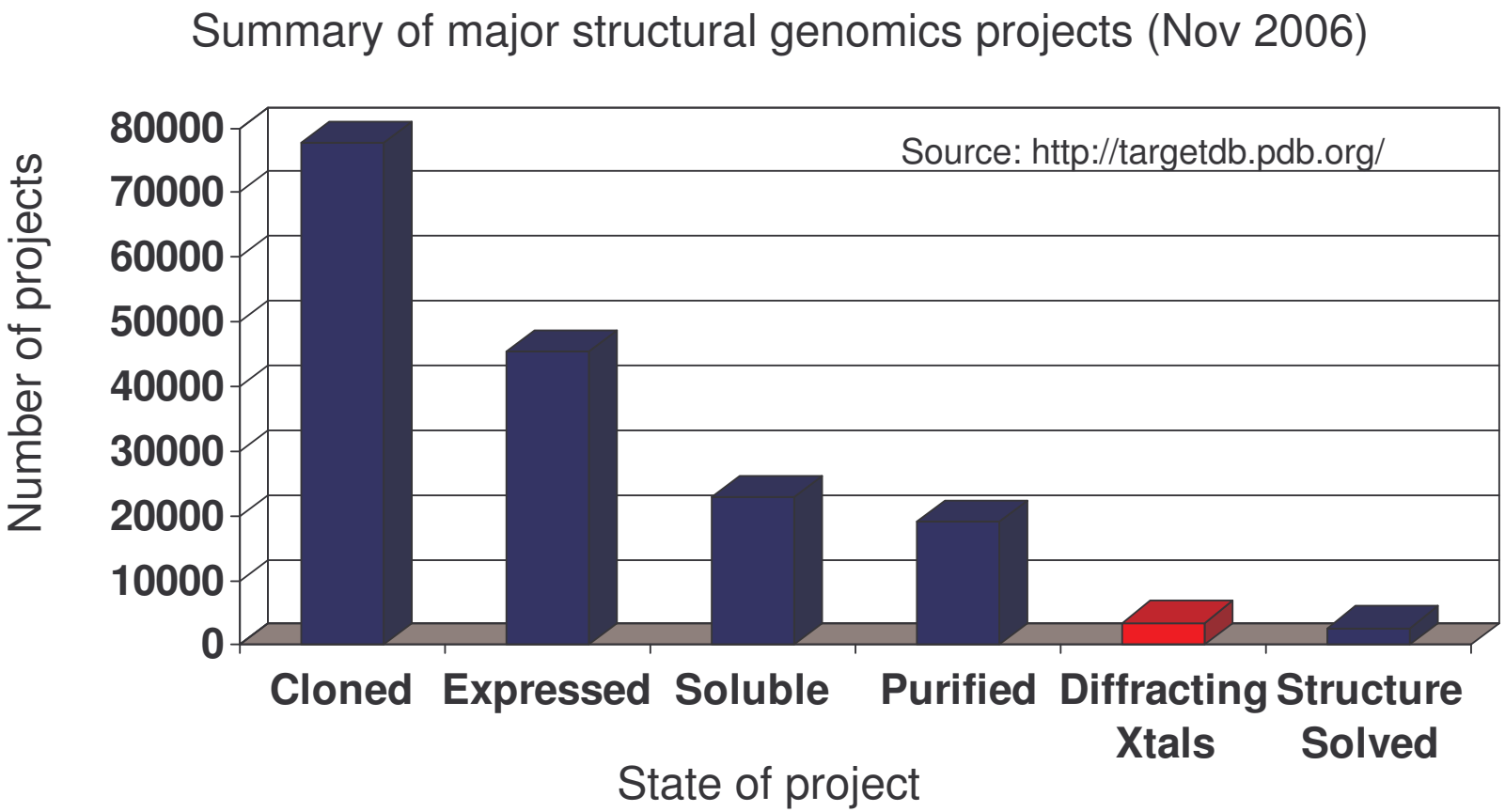
Protein Crystallization: An Introduction

Simplify Complexity

Agenda

1. Protein Crystallization: An overview
2. Protein requirements
3. Crystallization: Theory
4. Crystallization: Techniques
5. Crystal analysis and structure determination

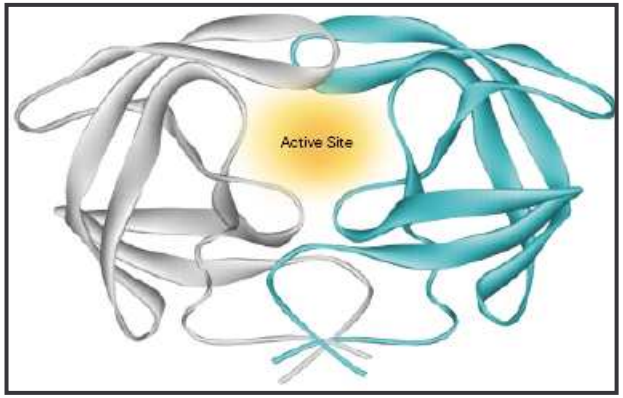
Crystallization is still the bottleneck of structure determination



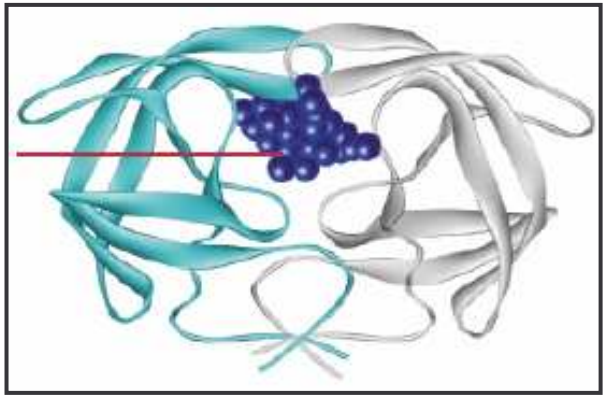
There is still a lot of work to do… let‘s start crystallization!

Why protein crystallization ?

- To determine a protein structure
- To understand the protein's binding to a ligand / substrate / drug
- To understand the protein's enzymatic mechanism
- Method: X-ray crystallography
  - No limits in size, whole viruses and ribosomes have been analyzed
  - crystals serve to amplify the signal (one crystal contains 10^9 to 10^{13} molecules)
  - Used to determine the exact position of atoms of a protein
  - Used for co-crystallization with ligands
  - 80% of all protein structures are determined with X-ray (rest using NMR)

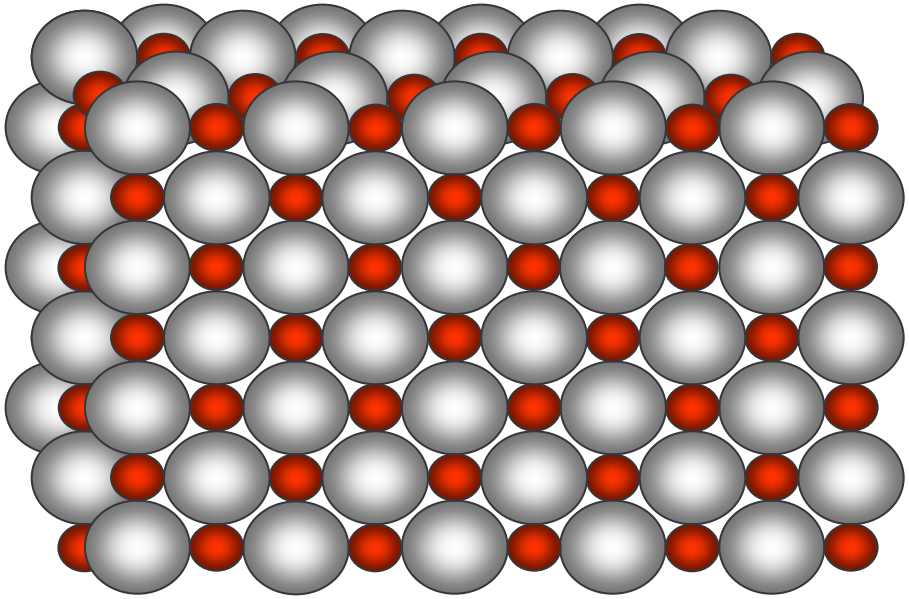


HIV Protease



HIV Protease with Drug

What is a crystal?

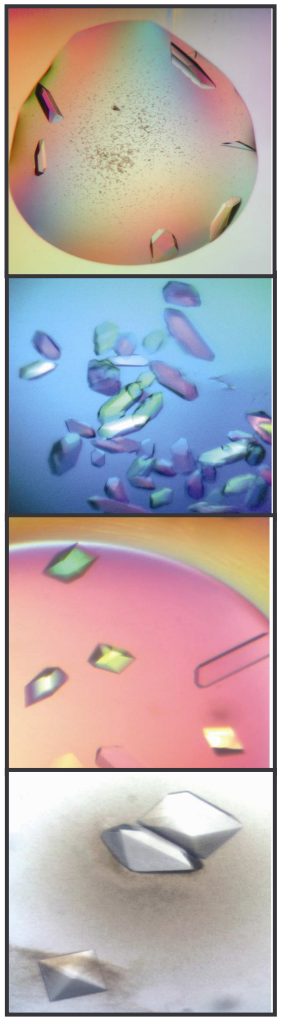


A crystal is a three-dimensional, periodic assembly of individual molecules that align themselves in a repeating series of "unit cells".

This is true for small molecule (e.g. salt) crystals and large molecule (protein) crystals.

Crystallography (from the Greek words *crystallon* = cold drop / frozen drop, and *graphein* = write) is the experimental science of determining the arrangement of atoms in solids. (source: Wikipedia)

What are the features of a protein crystal?



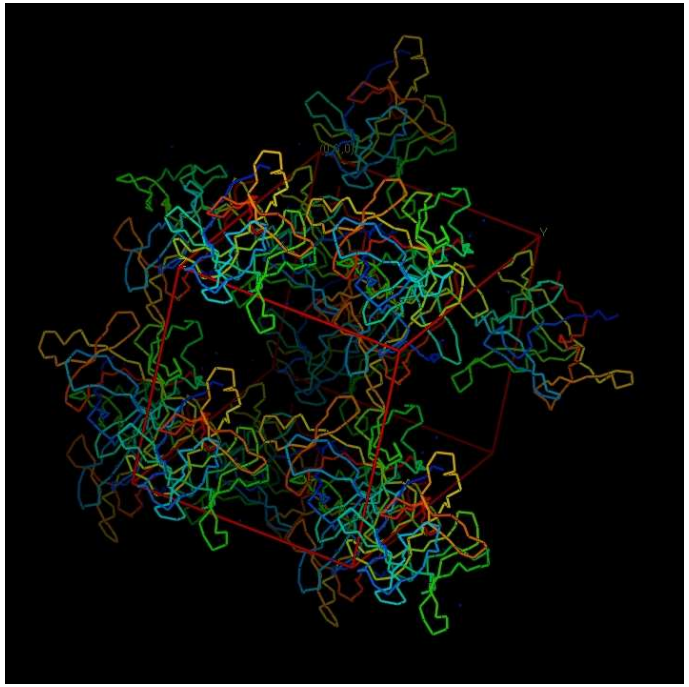
In contrast to salt molecules, which are tightly packed, protein molecules are

- Larger (1000 and more atoms vs. 2-10 atoms)
- Irregular in shape
- Have an irregular surface

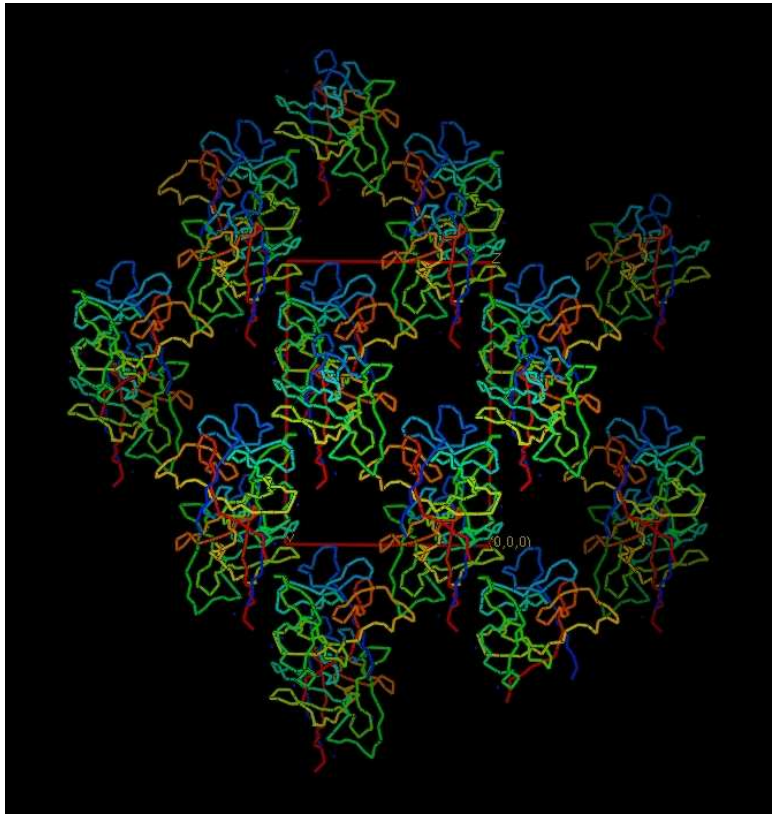
Therefore, protein crystals

- have less contact area between individual molecules
- are softer and more fragile
- The same protein can grow into differently shaped crystals, using different intermolecule contact areas
- Have special requirements to pH, salts, temperature to stay in native conformation
- Show weaker birefringence (rainbow color in polarized light) than salt crystals
- Contain around 50% water that forms channels
- Should not dry out & be handled with care
- Are a lot more difficult to grow than salt crystals

Protein crystal model images (Kubicek et al., 2007)

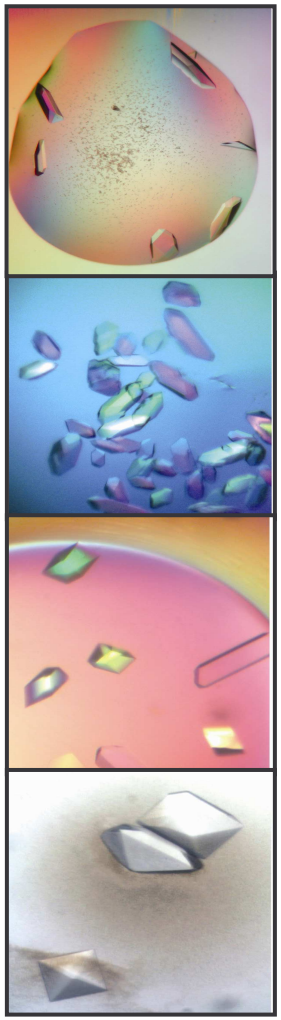


Protein unit cell of interleukin-beta crystal
See unit cell in red



Same crystal. See small contact areas between protein molecules and large channels filled with water/buffer

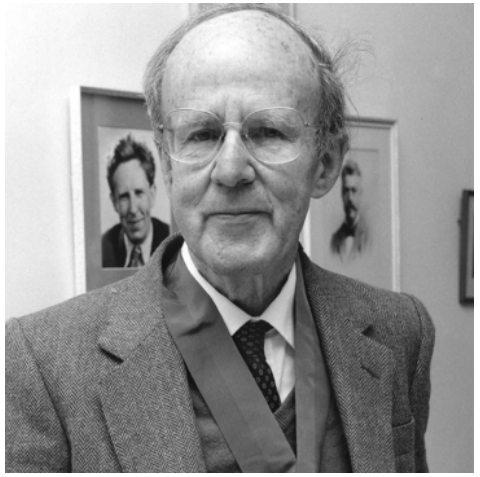
History



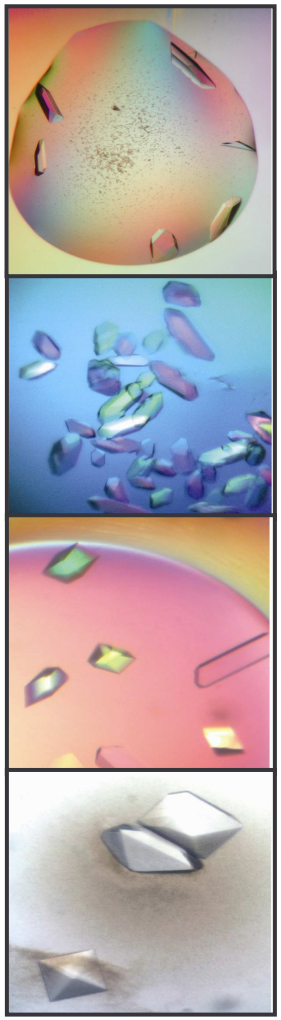
- The very first protein crystal (phycoerythrine, a pigment-containing protein from cyanobacteria) was described in 1894, even before x-rays were known
- In 1912, only 4 years after x-rays were used for medical purposes, Max von Laue, P. Knipping and W. Friedrich observed for the first time the diffraction of X-rays by crystals. This discovery, along with the early works of P. Ewald, W.H. Bragg and W.L. Bragg gave birth to the field of X-ray crystallography (of salt crystals, for a start)
- First protein structure determined was myoglobine in the late 1950s by Perutz & Kendrew
- Today, > 40.000 protein structures are deposited in the protein database (PDB), including whole viruses and ribosomes

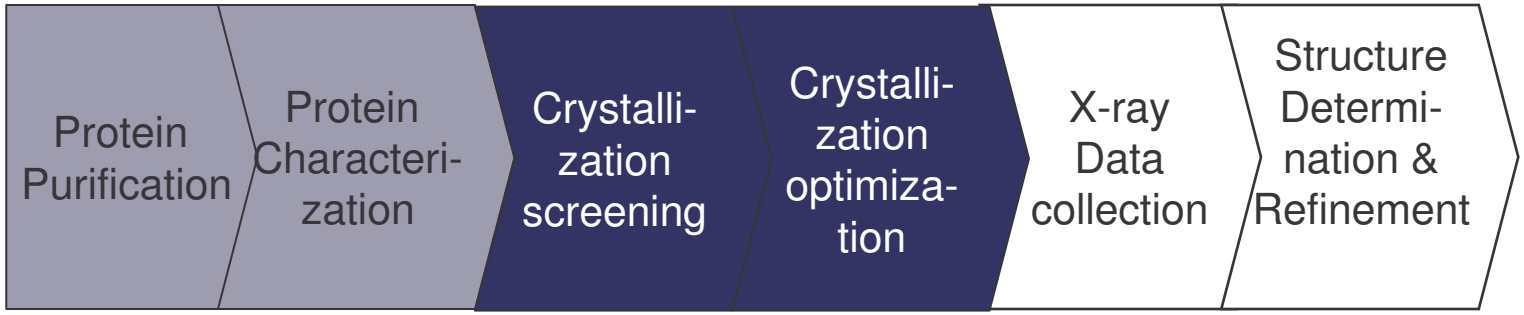
« The X-ray study of proteins is sometimes regarded as an abstruse subject comprehensible only to specialists, but the basic ideas underlying our work are so simple that some physicists find them boring. »

Max F. Perutz, Nobel Prize Chemistry, 1962



Protein structure determination workflow



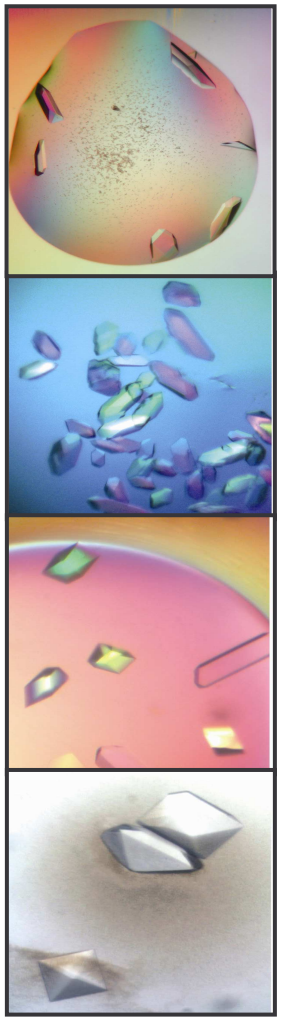


Usually, you as biochemists will be taking care of the „blue areas“, protein crystallographers will help you doing the X-ray analysis

Agenda

1. Protein Crystallization: An overview
2. Protein requirements
3. Crystallization: Theory
4. Crystallization: Techniques
5. Crystal analysis and structure determination

Proteins need to be pure for crystallization

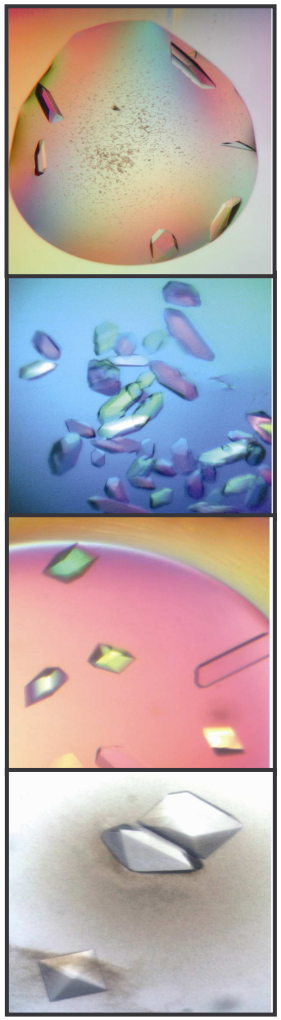


In order to form a highly ordered, homogeneous crystal lattice, proteins have to be

- Pure = not containing any contaminations
- Correctly folded
- Homogeneous = not containing any truncated proteins, isomorphs, incorrectly or differently folded molecules
- Proteins are flexible, have multiple conformations, but in the crystal, you have typically only one conformation -> so everything that is flexible will make crystallization difficult
- Not aggregated
- When a ligand is bound, it should be added in excess so that all protein molecules have a ligand bound to them
- If small tags are used for purification, they need to stay attached or removed completely but the protein solution should not contain a mixture

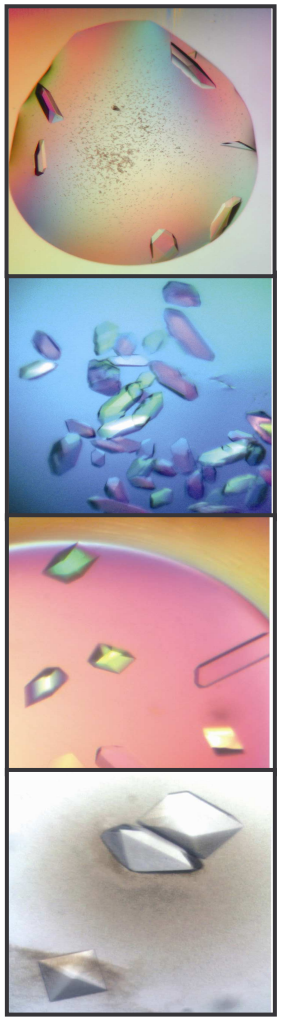
When you have an unsuccessful crystallization experiment, check purity and homogeneity first!!

How to evaluate a protein before crystallization



- SDS-PAGE: overall estimation of protein size and purity
  - Can show artifacts for very basic or acidic, glycosylated and membrane proteins
  - Resolution to several kDaltons, e.g. difference of 6 histidines (removal of tag) can be detected, not precise but cheap
- Size-exclusion chromatography: Size, Purity, Homogeneity
  - Precision depending on resolution of column (e.g. 50 or 150 kDa), can at the same time be used as final purifcation step
- Circular Dichroism (CD) spectroscopy: Folding
  - See secondary structure (percentage beta-sheet, alpha helix) to check for correct folding
- Mass Spectrometry (MS): Size, purity, modification
  - Very precise measurement of molecular weight in daltons, also shows glycosylation, phosphorylation
- Dynamic Light Scattering (DLS): size, and aggregation status= size distribution
  - E.g. discriminate between protein monomers (Ø 3 nm or 25 kDa) and aggregates (Ø 20 nm = 500 kDa)

How to handle a protein

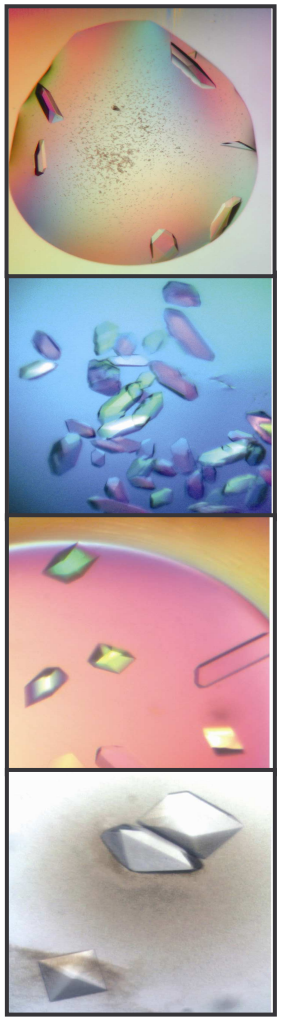


- Have a reproducible source of protein, a stable expression + purification that allows you to start with several mg of protein at least
- Choose the right source for your protein:
  - E.g. if it needs posttranslational modifications, express it in a eucaryotic system (insect cell, mammalian…)
  - E.g. if it is temperature-sensitive, why not try a homologue from thermophilic bacteria
- Always use fresh protein for crystallization (do not freeze if possible)
- Do not mix different purification batches (-> homogeneity)
- Find out what your protein likes…
  - pH & buffer substances
  - Salts, detergents
  - Temperature
  - Cofactors / ligands
- ….But: When you set up your crystallization experiment, consider which additives your protein solution already contains !

Agenda

1. Protein Crystallization: An overview
2. Protein requirements
3. Crystallization theory
4. Crystallization techniques
5. Crystal analysis and structure determination

Protein Solubility…



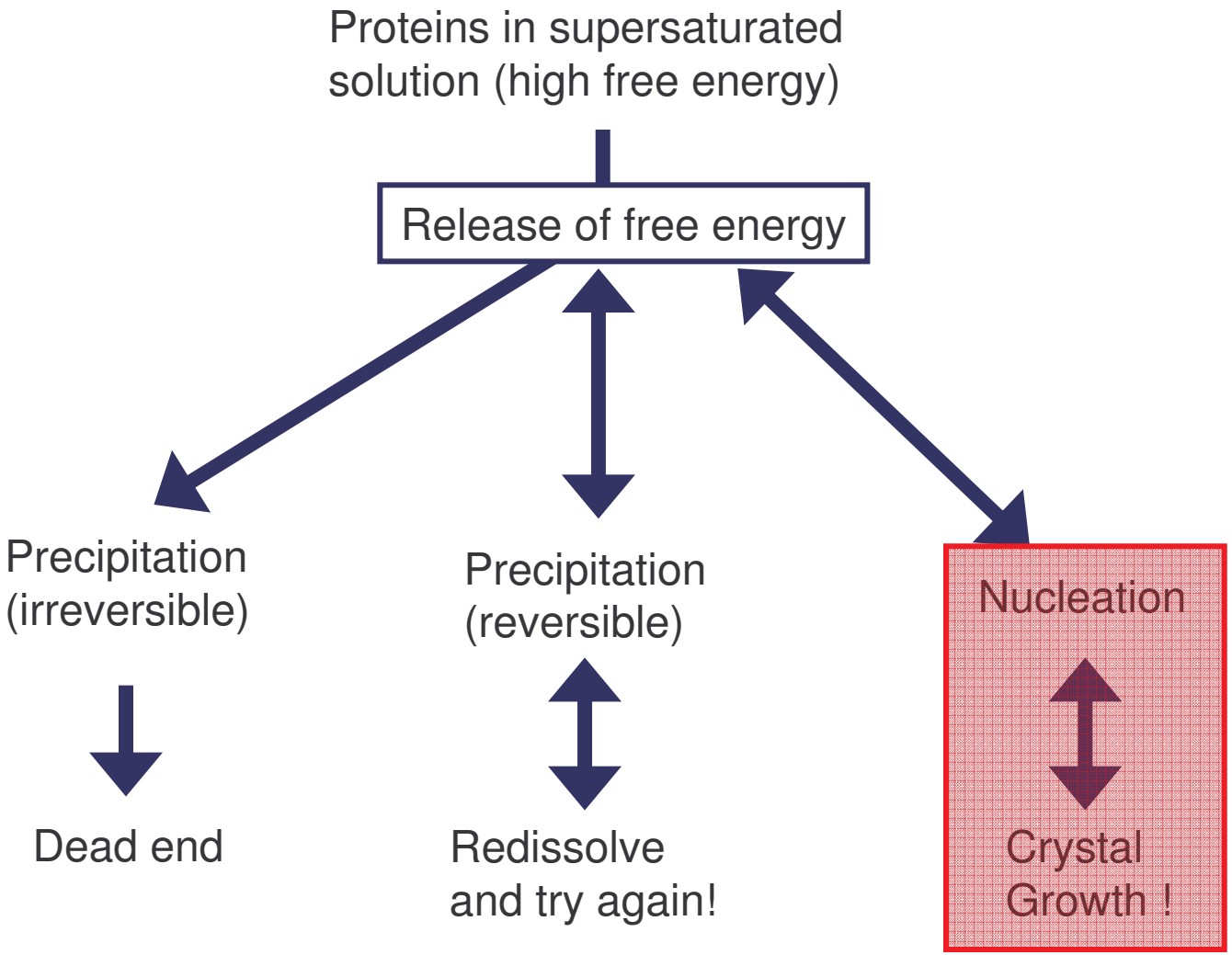
Protein solubility depends on…

- the protein itself, its surface charges and structure
- the solution surrounding the protein (pH, salts, organics…)
- temperature
- the protein concentration

Overall, the solubility of a protein is hard to predict, and can only be found empirically.

However, supersaturated protein is needed for a successful crystallization experiment!

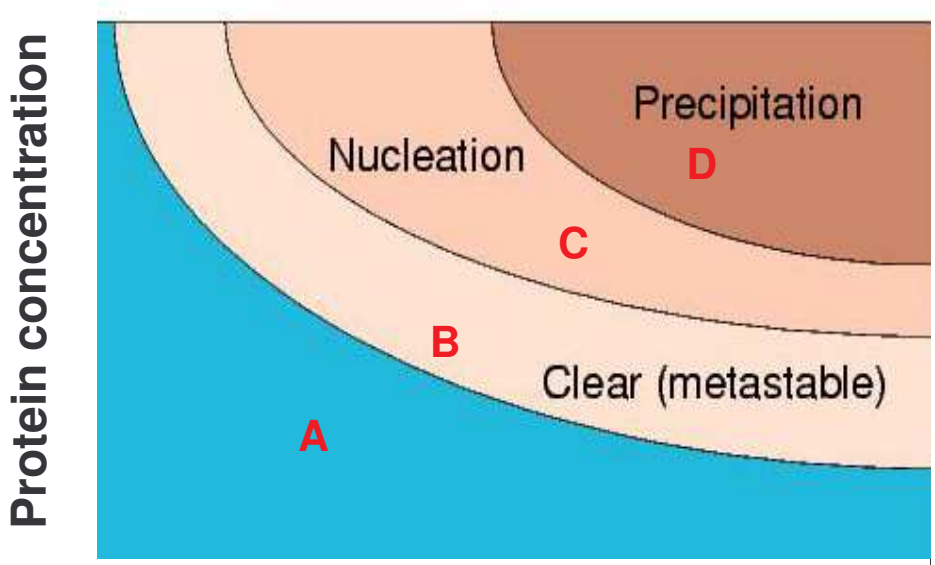
What can happen to a supersaturated solution?



The Phase Diagram

Undersaturation Zone (blue)
A protein and crystallizing agent concentration too low; no crystallization

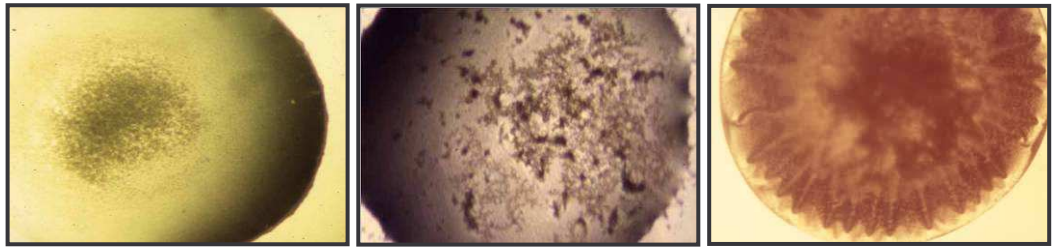
Supersaturation Zone (brown)
D Precipitation: concentrations too high, precipitates appear
C Nucleation: critical nuclei formation
B Metastable: crystal growth, if nucleation has happened before



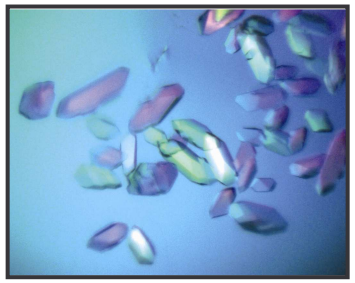
⇒First goal is to find the nucleation zone! Note that the size of nucleation zone is exaggerated here which makes it harder to find than you would think

What does my experiment look like?

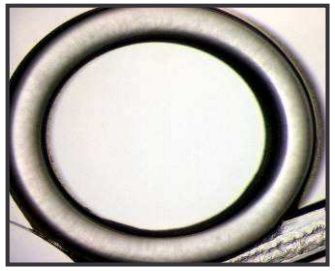
D

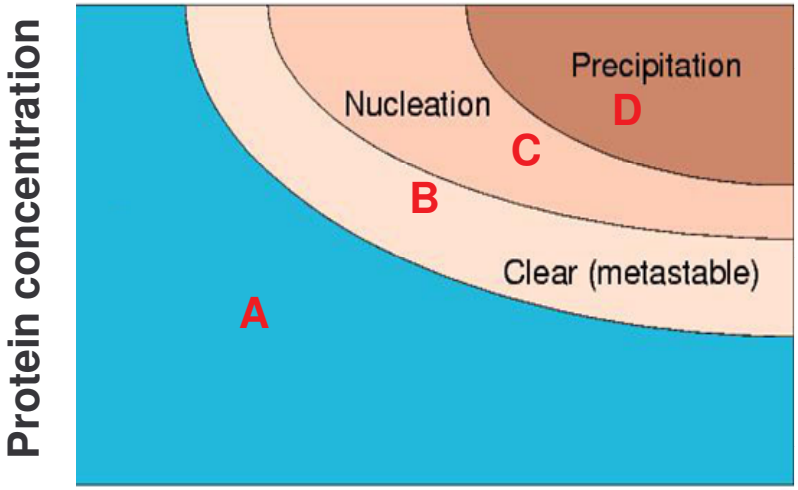


B - C



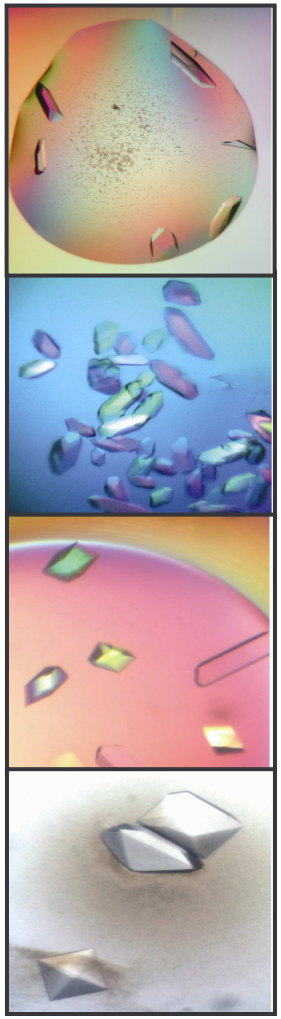
A





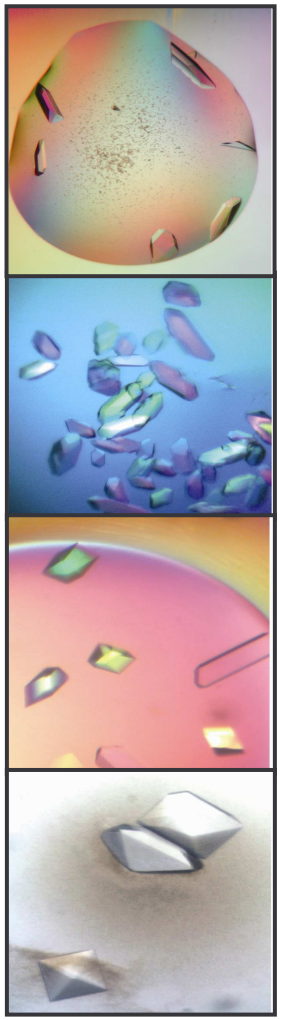
Crystallization will occur between the nucleation and clear metastable zone

Why is crystallization still the bottleneck of structure analyis?



- Finding the right crystallization conditions is a tedious try-and-error process
- Usually many hundreds or thousands of conditions are screened
- Data interpretation is difficult – crystals and precipitates come in many different shapes
- Most researchers use conditions that have been shown successful for other proteins, but this represents only a very limited portion of the chemical space
- First attempts to a logic approach:
  - Walk where no man has gone before… through the chemical space and make variations (grid screens) wherever you see first good signs, thereby using completely new conditions
  - Walk through the phase diagram: Look at all your wells, see if you can reverse precipitation, or grow crystals from clear drops
  - Know as much as possible about your protein: Some just need addition of the suitable ligand, ion or detergent to stabilize the good conformation that will grow into crystals

How to choose the right crystallization conditions?



This is still a try-and error process. We recommend the following strategy for a systematic approach:

1. Determine the optimal protein concentration for each class of precipitant (salts, organics, PEGs) using the pre-screening assay

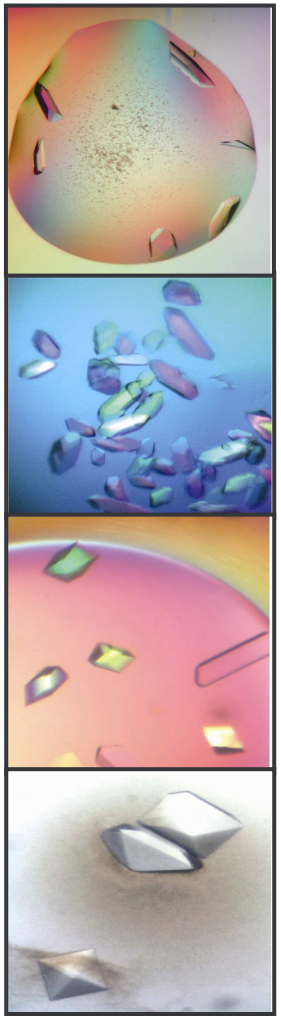
2. Perform an initial screening with broad-range screens. Combine a sparse-matrix and grid screen approach (e.g. JCSG+ / PACT screen)

3. When promising precipitants have been identified, perform a more detailed screening using this precipitant (e.g. AmSO4 or PEG)

4. When you obtain a hit, optimize it with a grid around the hit conditions (vary salt, pH) (OptiSalt)

Strategic approach will get you to crystals faster

What does the phase diagram tell us?



- Nucleation (initial start of crystallization) happens at a higher saturation than crystal growth!
- As protein is used up to form the crystal, the concentration drops, and you get to clear/metastable zone where crystal growth continues to form nice large crystals
- If the experiment happens too deep in nucleation zone, many nuclei will form, leading to a „shower“ of crystals
- If the experiments happens too close to the border of nucleation / metastable phase, crystal growth stops due to depletion of protein, and you only get small crystals.

⇒ **Many researchers today try to get in control of the nucleation and crystal growth process instead of following a try-and-error procedure**

⇒ **However, this is not easy as the zones of the phase diagram vary between proteins and between protein / precipitant combinations**

Good vs. Bad crystal

Some crystals look very nice, but do not diffract the X-ray beam.

These crystals look ugly, but they diffract to 1.6 Å.

Why is that so?
-> It's the internal order that counts, not the shape

Pictures courtesy of Terese Bergfors, Uppsala University. See her online tutorial at http://xray.bmc.uu.se/terese/

Agenda

1. Protein Crystallization: An overview
2. Protein requirements
3. Crystallization theory
4. Crystallization techniques
5. Crystal analysis and structure determination

How to reach the target = nucleation zone?

You need to reach supersaturation

- By concentrating the protein (y-axis of the phase diagram)
- By adding different concentrations of a precipitant (x-axis)

Precipitants are defined as chemicals that lower the solubility of a protein in solution. These can be:

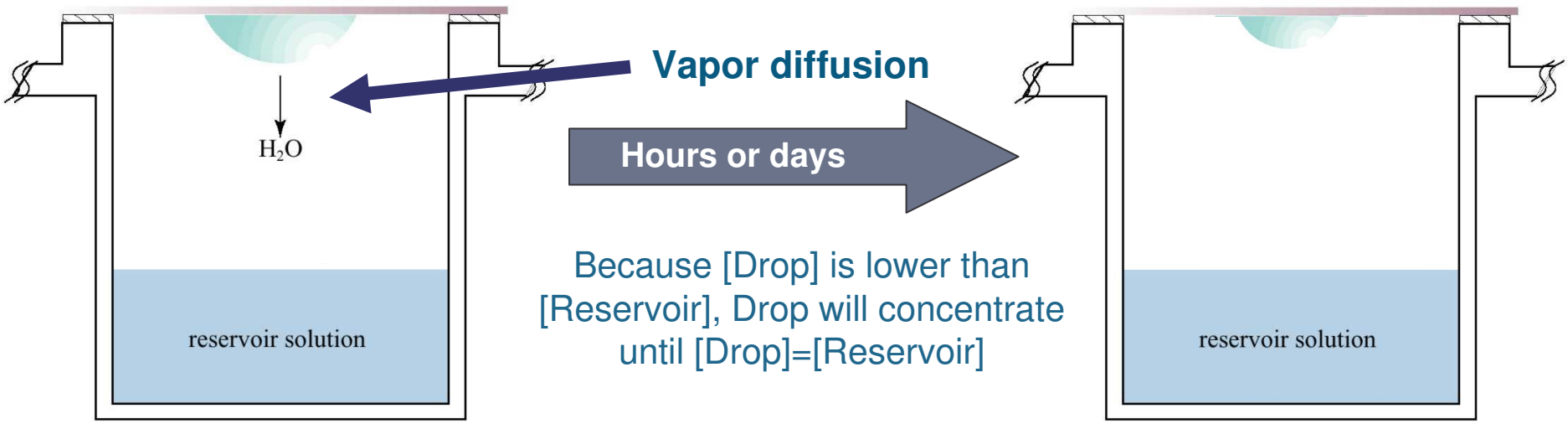
- Salts (e.g. $AmSO_4$, NaCl) „salting out“ process
  - sometimes salts also help crystallization by binding specifically to the protein
- Organic solvents (e.g. ethanol, MPD…)
- PEGs (of different molecular weights)

Commonly used techniques use concentration and / or precipitation:

- Vapor diffusion
- Microdialysis / microbatch
- Counterdiffusion

Most popular technique is vapor diffusion because you have many ways to control your experiment (start, end point, travel back through the phase diagram)

General Techniques: 1. Vapor diffusion

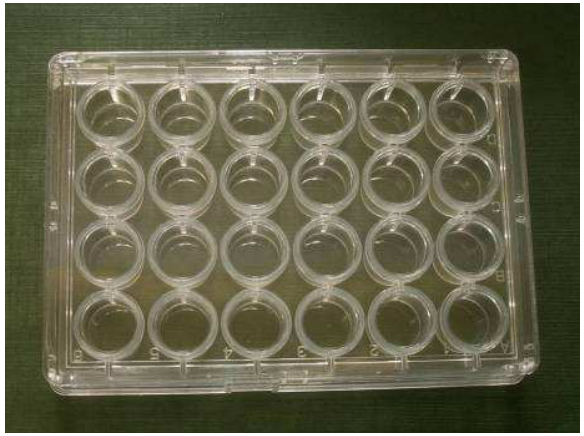


Setup:

- Mix e.g. 2 µl of protein solution with 2 µl of precipitant solution
- Over time, water will evaporate from the drop, increasing the concentration of both protein and precipitant
- When you have the right concentration and combination of chemicals, pH and temperature, crystals will appear in the drop!

Vapor diffusion setup methods

A



A. Conventional setup using 24 well cell culture plate (Linbro), siliconized glass slides and grease

B

B. Next generation setup:
- EasyXtal Tools with grease-free, screw on caps.
- Easy manipulation of reservoir solution: Just open and close the screw-caps =crystallization supports.
- Also available pre-filled

Vapor diffusion: Hanging vs. sitting drop

Hanging
24 wells

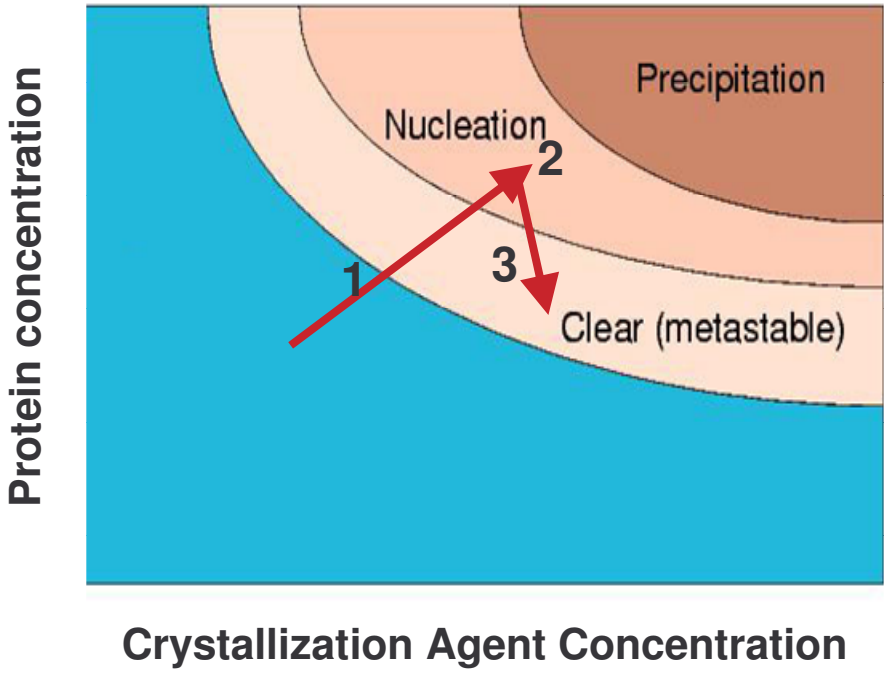
Sitting
96 wells

	Hanging	Sitting	
Imaging	+++	+	
Crystal recovery	+++	+	Easier access x-tals tend to stick in sitting
Crystal quality	+++	++	Some customers think it is better
Sealing	+++	+	Tape is less efficient
Ease of setup	+	++	Can use multi-channels
Automation friendly		+++	
Proactive screening	+++		Saves protein and time

Vapor diffusion and the phase diagram

1. Protein concentration and crystalllizing agent concentration both increase through water evaporation
2. Nucleation (initial start of crystallization) happens at a higher saturation than crystal growth!

3. As protein is used up to form the crystal, the concentration drops, and you get to clear/metastable zone where crystal growth continues



-Changing the reservoir solution (decreasing or increasing concentration) allows you to influence your experiment at all times
-This way, you can travel across the phase diagram (pro-active strategy)

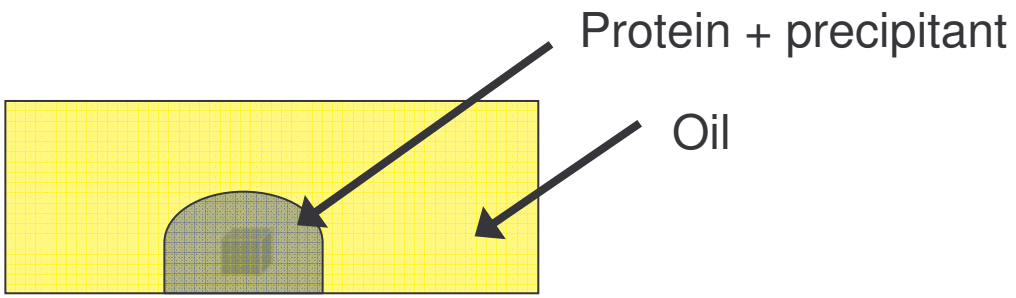
General techniques 2.: Microbatch using oil

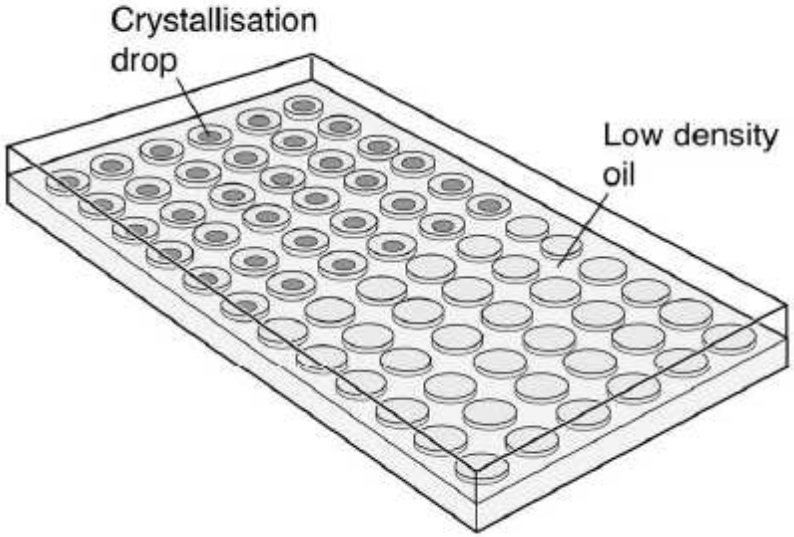
1. Protein + precipitant are mixed and the drop immersed in oil

2. Supersaturation is reached immediately

3. depending on the nature of the oil, evaporation can or cannot occur (example: silicon oil is more air permeable than paraffin oil, often mixtures of these two are used)

4. If evaporation occurs, you can start in stable part of phase diagramme and move diagonally up, but you do not have an endpoint like in vapor diffusion

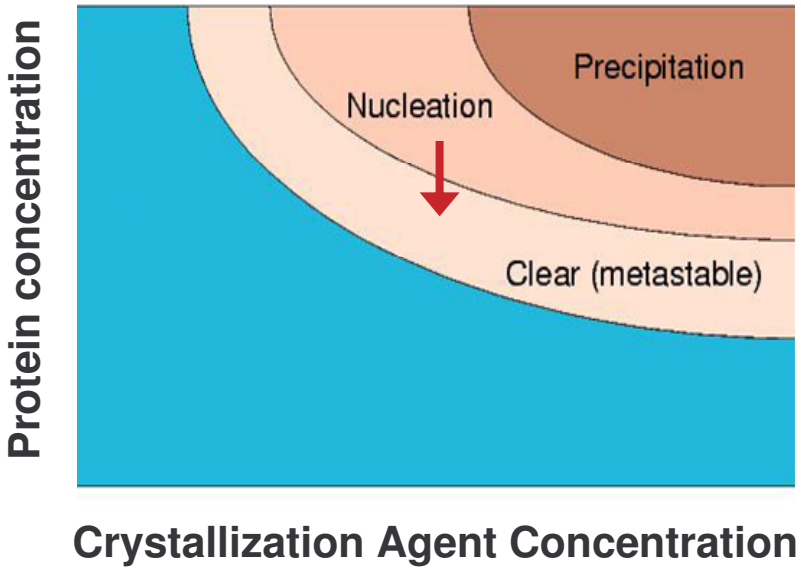




Microbatch and the phase diagram

1. You start in supersaturated zone (needs to be determined before)

2. Protein concentration decreases as crystal grows



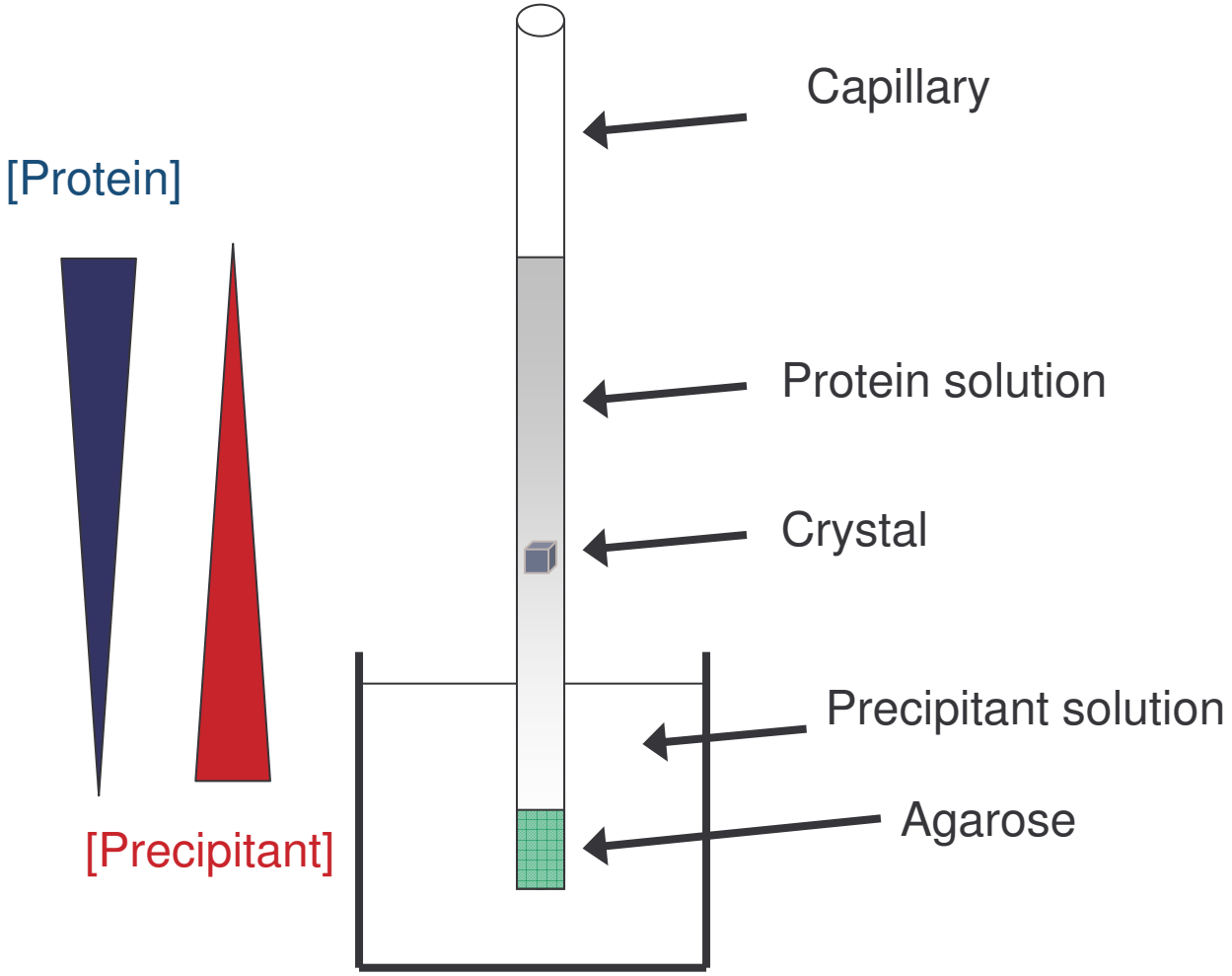
You need to know a lot about the solubility of your protein in advance to set up your experiment at the right concentration – and not hit precipitation or undersaturated phase

More flexibility if oil allows for evaporation
Once experiment is set up, it is quite difficult to manipulate

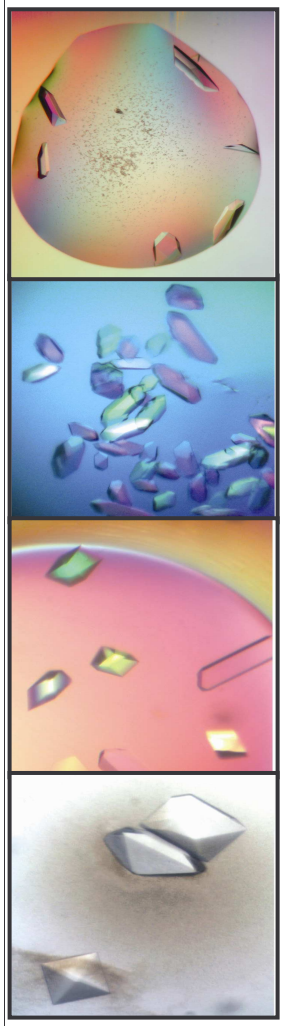
General techniques 3. Counterdiffusion

Protein and precipitant diffuse along the capillary, generating many microconditions, one of them maybe leading to crystal formation

In each region of the capillary, individual conditions form that have their own phase diagram



Advanced techniques: 1. Seeding



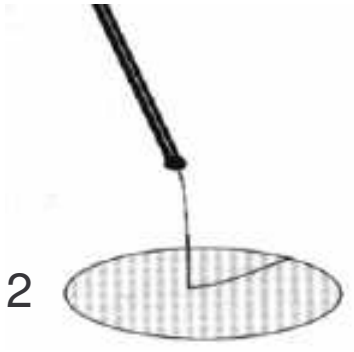
What is seeding?

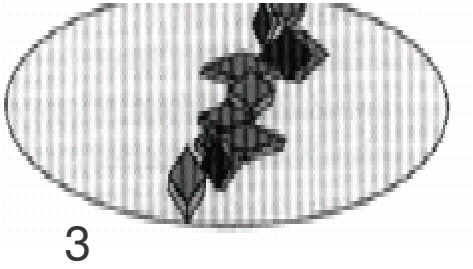
- Optimization method
- A. Homogeneous seeding: Crystallization method where a crystal of the same protein is used as nucleant in a metastable protein solution
  - Microseeding: small crystal or small bits are used
  - Macroseeding: larger crystal is used
- B. Heterogeneous or Cross- seeding: A small crystal of a similar or even different protein is used as nucleant

When to use seeding?

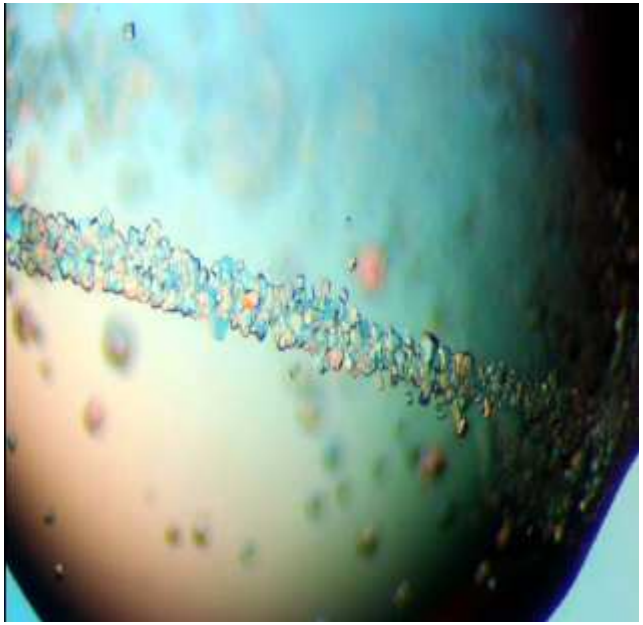
- When you do not get crystals at all (meaning the nucleation zone is very small and hard to find) => heterogeneous seeding
- When crystal growth stops because protein concentration gets too low, or if growing areas are blocked by proteins in different conformation. Wash your crystal and put it in a fresh supersaturated solution, then it can continue to grow => homogeneous seeding

Streak Seeding (E. Stura, 1992)

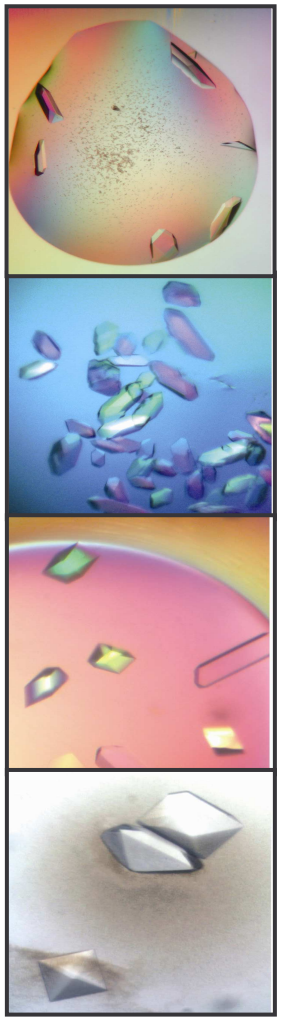




1. Pick up microcrystals with seeding tool (e.g. cat whisker)
2. Seed in supersaturated protein + precipitant solution
3. Crystals will grow along the streak



Advanced techniques: 2. Soaking



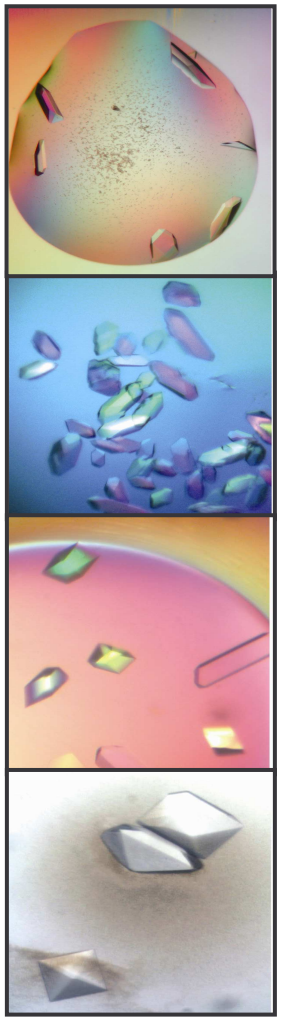
What is soaking?

- Crystal is taken out of the mother liquor and soaked in a solution containing
  - Cryoprotectant (e.g. glycerol) or
  - Ligand /substrate or
  - Heavy atom solution

Why soaking?

- To protect crystal from damage in the x-ray beam (replace water by cryoprotectant) or
- To determine the binding pocket of a ligand or substrate in the active site of an enzyme or
- To produce heavy atom derivatives of the protein crystal, using the MIR method (multiple isomorphous replacement)

Soaking



What can go wrong?

- It can happen that the binding of the ligand / substrate / heavy atoms influences the folding of the protein and its 3D structure

  => the interaction areas between the protein molecules change, and the crystal dissolves

What to do?

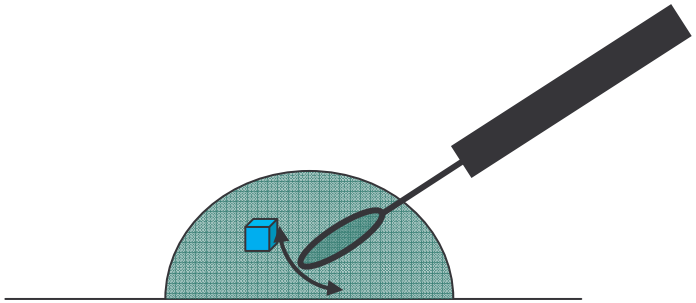
- Co-crystallize your protein with the ligand
- for heavy atoms: use an expression system where you add modified amino acids (e.g. selenomethionine) to the medium or IVT assay

Crystal harvesting

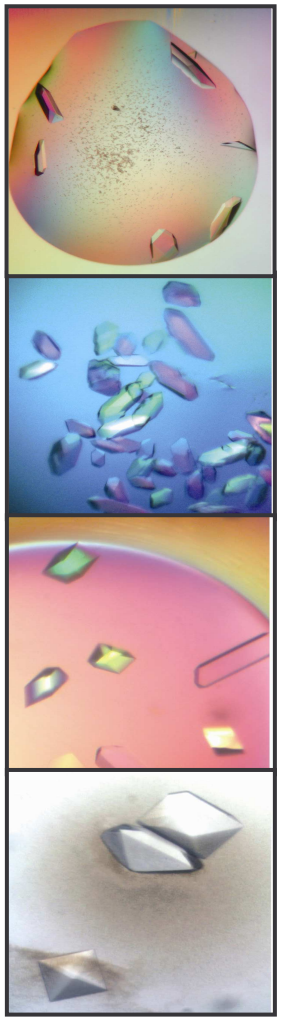
1. Harvest crystal from the crystallization drop using a small nylon loop or metal mesh

2. Soak crystal in cryoprotectant & freeze in liquid nitrogen

3. Use the loop to hold the crystal into the x-ray beam



Why cryoprotection?

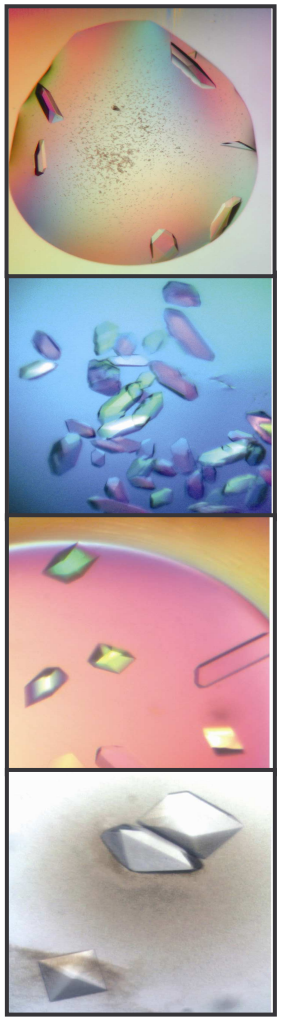


- Advantage of freezing:
  - X-ray radiation is very strong, can destroy the protein crystal
  - Damage caused by radicals and followed radical chain reaction across the crystal
  - At 100 K (liquid nitrogen temperature), radicals stay local („frozen“), reducing damage
  - Longer measurements can be done, at higher energy (synchrotron) to obtain more data / crystal
- Risk of freezing:
  - Large water channels can crack upon freezing, breaking the crystal
  - Cryoprotectants (glycerol…) are added to protect

Agenda

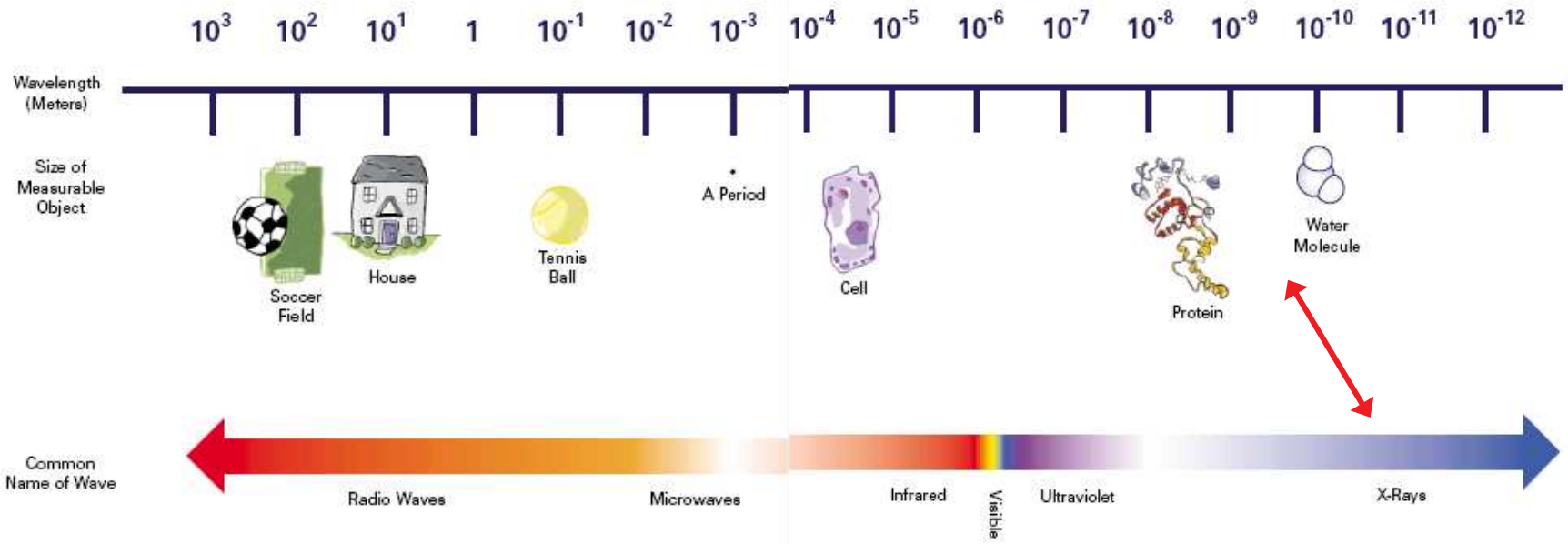
1. Protein Crystallization: An overview
2. Protein requirements
3. Crystallization theory
4. Crystallization techniques
5. Crystal analysis and structure determination

Why X-rays?



- When you want to see things with your naked eye, the resolution depends on the wavelength. Visible light has λ= 350-700 nm.
- In order to „see“ the atoms of a protein, you have to use „light“ that has the same wavelength of the smallest distance you want to see
- Distance between atoms in a molecule = 1.5 Å = 0.15 nm
- λ = 0.15 nm = x-ray radiation
- This radiation is not visible any more, but can be detected using a CCD detector (similar to a digital camera)
- For X-ray detection, you cannot use an optical lens, to obtain a real „picture“, you do mathematical calculations (Fourier transformation)

Sizes and wavelengths



Proteins are so small that X-ray light is needed to „see“ them

Sources of X-ray radiation

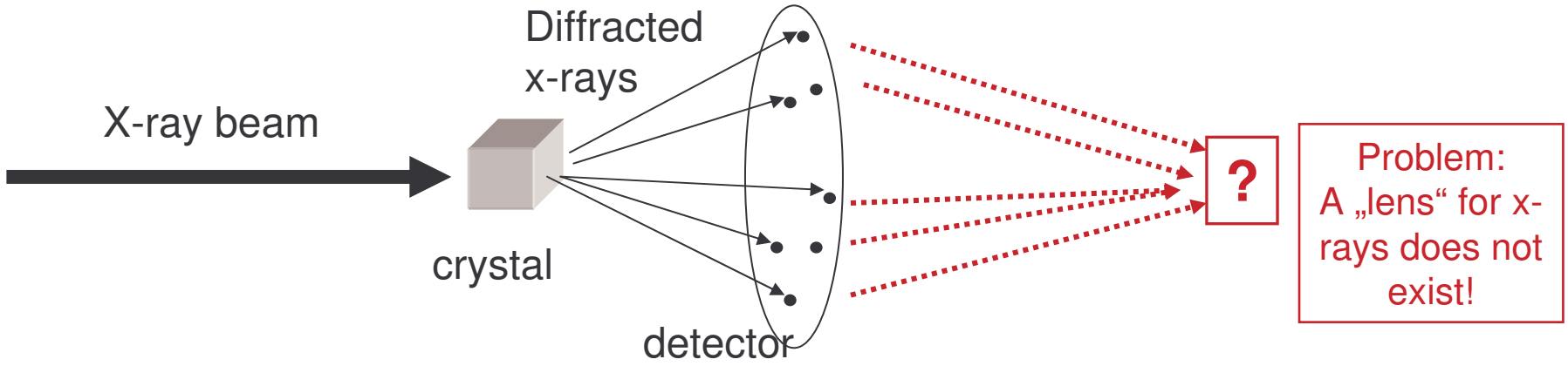
Rotating anodes

- Accelerate a beam of electrons to generate x-rays
-Often used in labs to check crystals
-Only one wavelength, depending on metal used for anode (typically Cu: 1.5 Å)

Synchrotrons

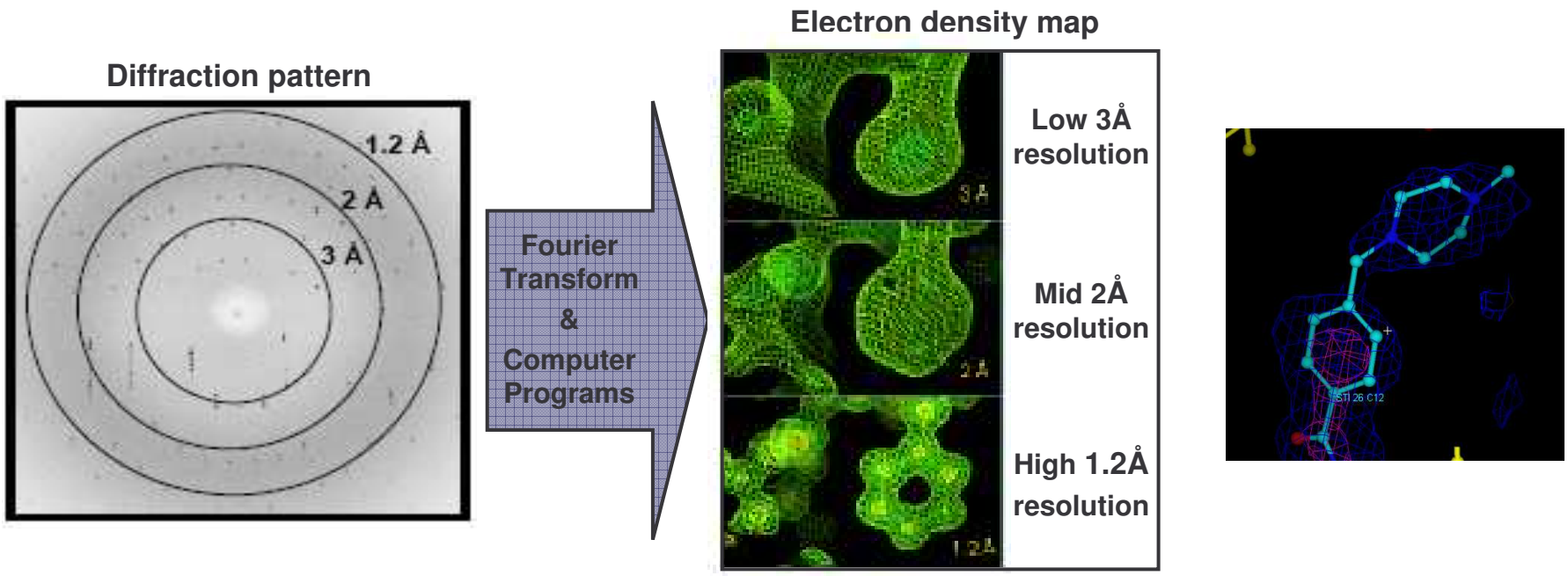
- Used to solve protein structure
- Large central service units
- Use magnets that bend the electrons
- Wavelength range, typically 0.5-3.0 Å
- Faster measurement due to up to 1000 fold higher energy

Diffraction



- When a crystal diffracts, the proteins‘ electrons absorb the energy of the X-ray radiation.
- When the excited electrons fall back to the ground state, they emit x-ray radiation which is now scattered
- When a large number of electrons in a crystal scatters the X-rays, interference occurs that can be additive or subtractive. Results of additive interference give rise to detectable signals
- The overall emission pattern of a crystal is a function of its 3-D structure, or, to be more precise, its electron densities (comparable to the ringing of a bell, the emitted sound waves are analyzed by the ear to recognize a sound)
- Crystal is turned in the beam 1-2° per image to get a full dataset

Resolution



What you can „see“ at different resolutions:

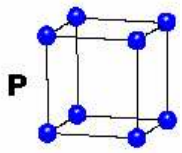
- 5 Å: shape of protein
- 3 Å: trace of polypeptide chain, amino acids can be fitted if sequence is known
- 2 Å: Peptide bonds and long side chains
-1 Å: Atoms

The higher the order within the crystal, the better the resolution!

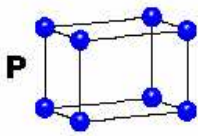
Space groups & unit cells

- A unit cell is the smallest entity of a crystal. It is defined by
  - The length of its a, b, c axes
  - The angles α, β, γ between the three axes

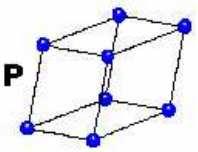
CUBIC
$a = b = c$
$\alpha = \beta = \gamma = 90°$



TETRAGONAL
$a = b \neq c$
$\alpha = \beta = \gamma = 90°$

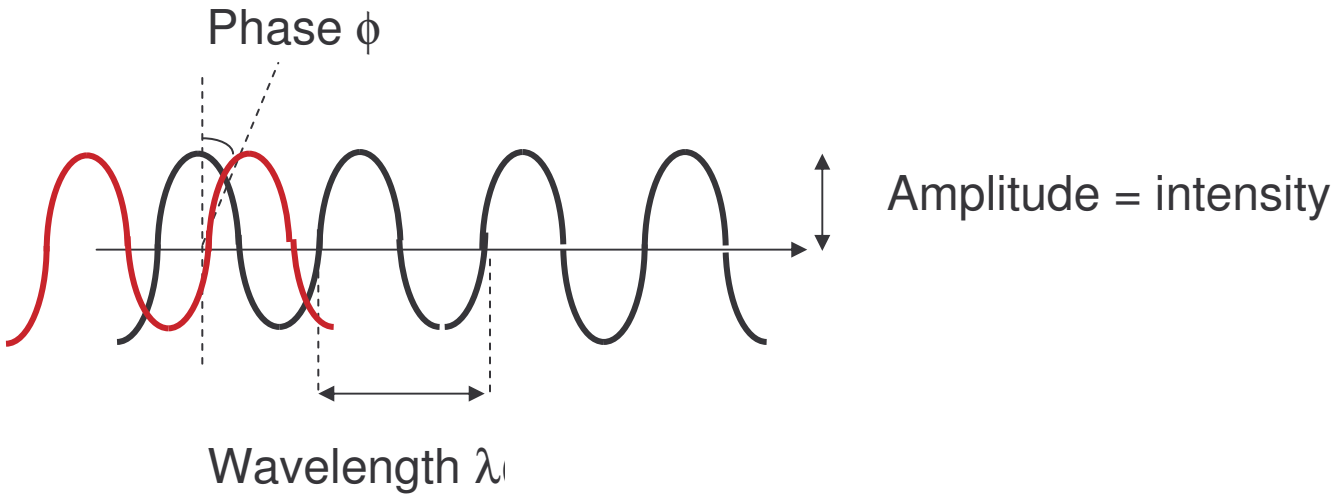


TRICLINIC
$a \neq b \neq c$
$\alpha \neq \beta \neq \gamma \neq 90°$



- The geometries of these building blocks define the symmetry of the crystal lattice: „space group“
- In proteins, 65 different space groups can be found. Rarely, you even see two space groups in one crystal (= twinning) which is very difficult to analyze
- The very same protein can grow crystals of different space groups, depending on crystallization conditions -> because precipitating chemicals mask certain charges on the protein surface, or help building a noncovalent „bridge“ between proteins

The Phase Problem…

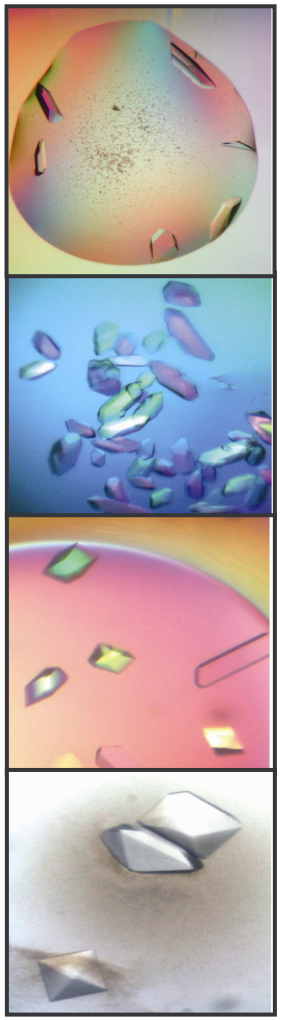


Three parameters are necessary to calculate
electron density from diffraction pattern data:

Wave identification:
H K L
1 1 1
1 1 2
1 1 3
2 1 1
2 1 2
1 2 1

1. Wavelength: does not change during measurement, is known
2. Amplitude: measured intensity, also known
3. „name“ of each wave: indicated by h,k,l numbers (see example)
4. Phase: direct measurement not possible

Good thing about the phase problem…

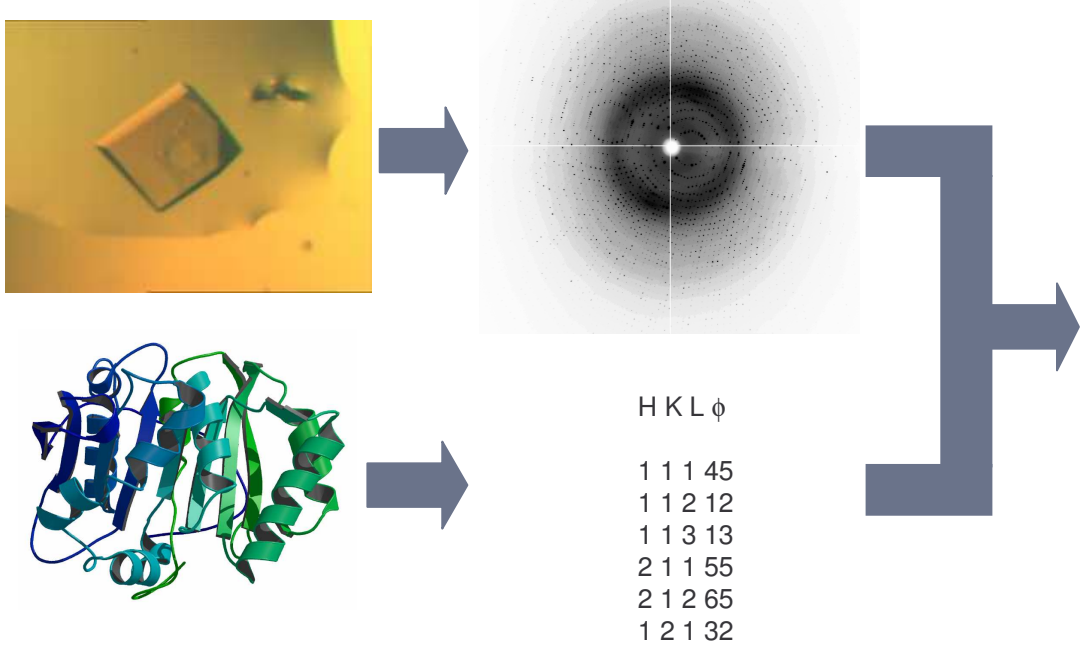


- You do not need to know all phases exactly, but a sufficiently „good guess“
- Bad phases do not show any recognizable features
- Good phases show you well defined electron densities (i.e. you can „see“ alpha helices). In this case, you know that your set of phases is a good enough guess to solve the structure
- Further improvement and refinement can be done by statistics and chemical knowledge

… and how to solve it: 1. Is it similar to something?

1. Molecular replacement:

- When the structure of a similar protein (e.g. homologue from different organism) was already solved, calculate the phase „backwards“ and use it as a good initial guess for the analysis of new protein
- => This is the easiest procedure, and our ability to use it obviously increases with the number of different protein folds solved and publicly available (structural genomics projects)



Picture Courtesy N. Tarbouriech, EMBL Grenoble

… or can we simplify it?: 2. Isomorphous replacement

- Soak crystal in solution containing heavy metals (e.g. $Hg^{2+,}$ $PtCl_{4)}$. Alternatively, selenomethionine can be added to the protein expression medium or IVT solution.
- The crystal should not change conformation, i.e. stay isomorphous.
- only few positions contribute to signal where heavy metals are incorporated (ideally: one position per unit cell). This pattern is a lot simpler than electron density of C atoms
- At the same time, X-ray diffraction patterns show higher intensities (amplitudes) because a lot more electrons contribute to the diffraction (80 in Hg compared to 6 in C)
- From a comparison of crystals with / without heavy atoms, you get a simpler signal only showing only contribution of Hg atoms.
- From this, the position of the Hg atoms in the unit cell can be calculated, and hence an estimation of the phase
- Phases of the "Hg structure" then give you a "good inital guess" of phases for your protein structure
- To refine the data, several different heavy metals are used, which absorb at different wavelengths (therefore measurement at a synchrotron is needed)

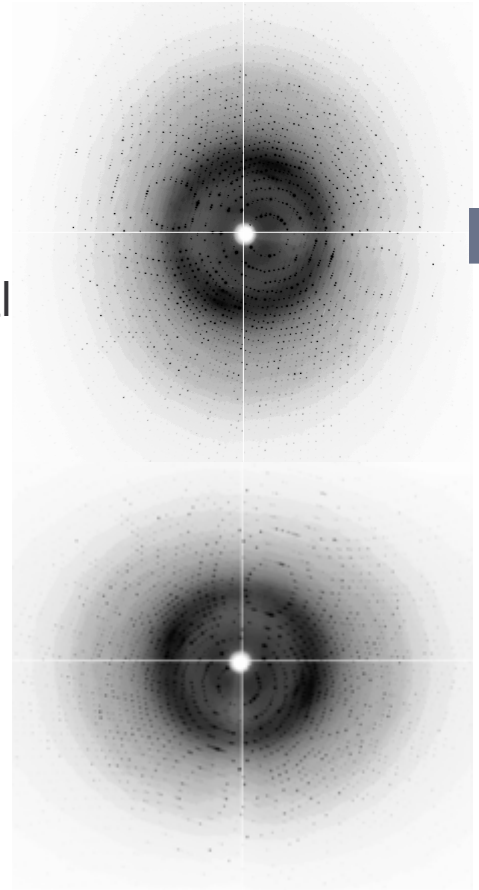
=> Multiple isomorphous replacement (MIR)

- Can be combined with MAD (multiple anomalous dispersion) where wavelenghts are used that excite the electrons at the inner shell of the heavy metals

Isomorphous replacement:

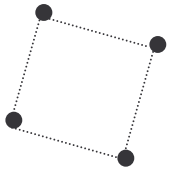
1.Native crystal

2.Crystal soaked in heavy atoms



3. X-ray analysis

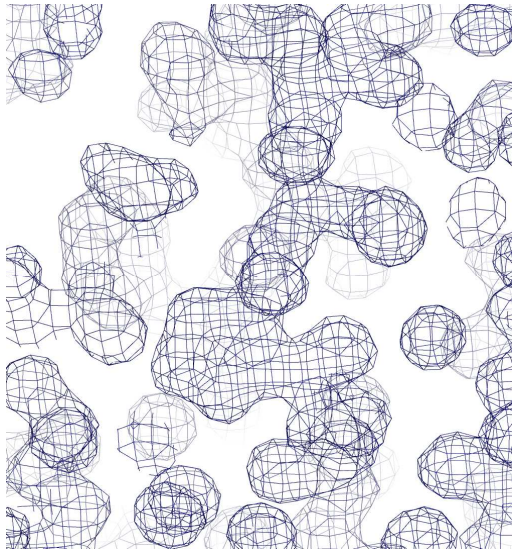
4.Difference +/- heavy atoms



H K L ϕ

1 1 1 45
1 1 2 12
1 1 3 13
2 1 1 55
2 1 2 65
1 2 1 32

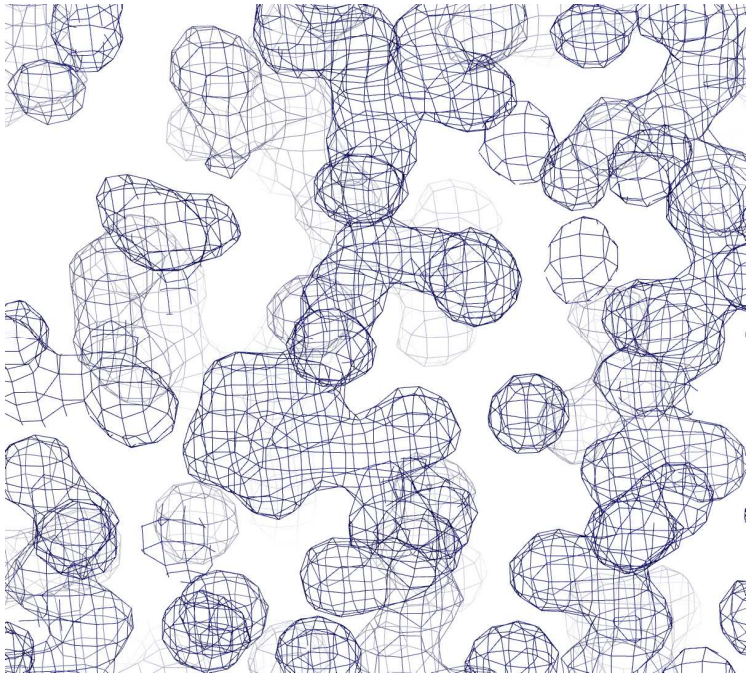
5.Calculate phases



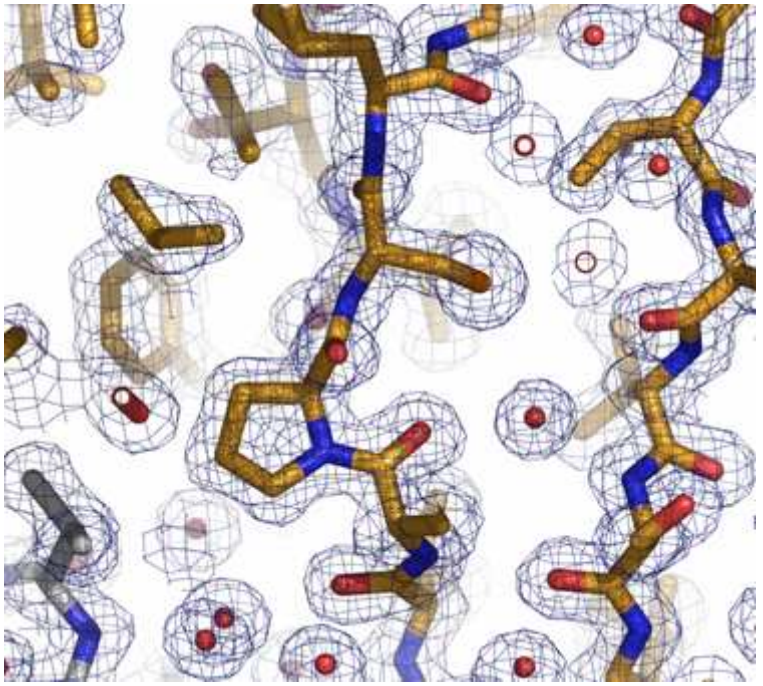
6. Determine electron density

Picture Courtesy N. Tarbouriech, EMBL Grenoble

Solving the structure

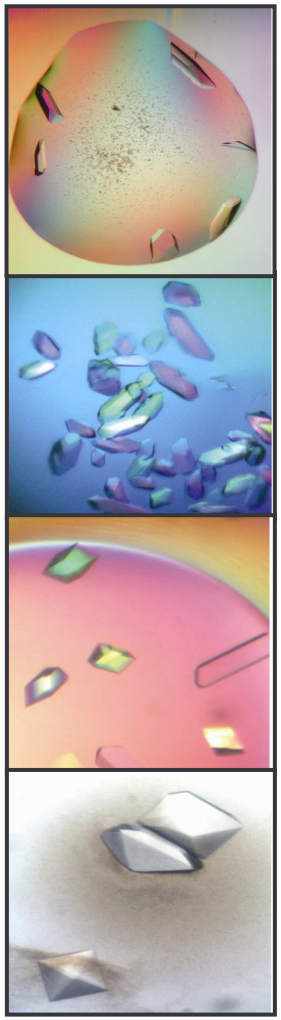


Electron density



Amino acid chain is fitted into electron densities via dedicated software

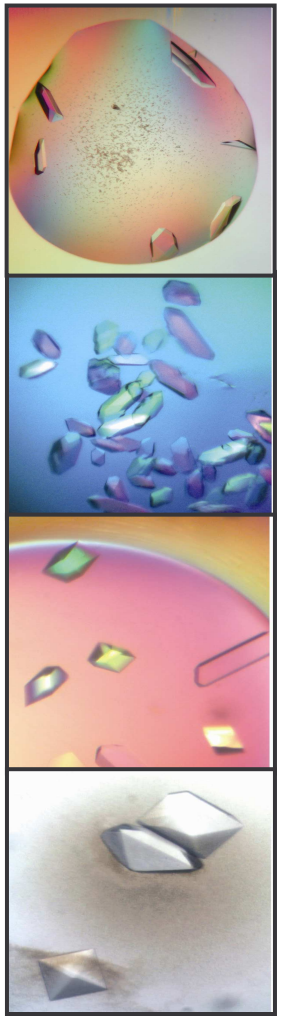
Structure refinement



First electron density maps are usually at lower resolution. Refinement can be done by:

- Measuring several crystals to minimize mistakes from flexible regions
- Compare with structure of similar proteins or domains
- Find the best phase for the experimental electron density
  - Try to fit amino acid sidechains, also by simulating movement of the side chains („virtual heating up“ of the molecule)
  - Re-calculate phase with optimized structure
  - Fine-tune structure obtained with recalculated phase
  - Etc. => Go through many cycles until structure is refined
- At the end: Compare your refined electron densities with original data, do they still fit? => determine possible error

Optimizing the crystal



- If you see that bioinformatics cannot help you to solve the structure of your crystal, it is necessary to go back to the protein crystal itself.
- To obtain a higher resolution, you have to increase tight packing of your protein by
  - Removing flexible loops
    - by PCR on DNA level or
    - protease treatment on protein level
  - Add inhibitor or ligand to block a defined conformation of your protein
  - Use only certain domains of your protein
  - Use a homologous protein from different organism, e.g. thermophilic bacteria
  - Optimize your crystal, using different crystallization conditions that might lead to different packing

Example for solved structure IL-beta (Kubicek et al.)

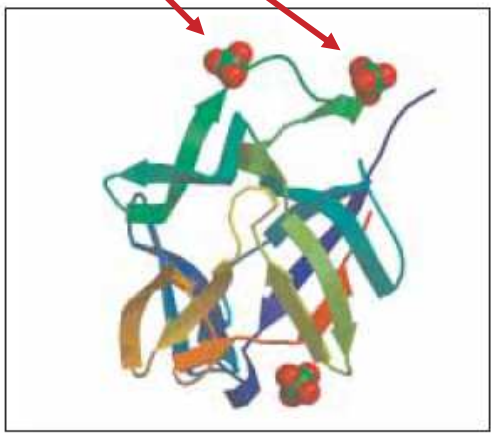
Interleukine-1-beta protein crystals were obtained with the QIAGEN AmSO4 Suite (2.0 M AmSO4, 0.1 M Na-acetate, pH 4.6, and protein at 28 mg/ml) in a vapor diffusion experiment. Size of crystal: 0,3x0.3x0.5 mm, diffraction 2 Å.

Size and shape of spacegroup

Sulfate from AmSO4 binds to protein, is part of crystal!

Cell	
Spacegroup	$P4_3$
a,b	54.6 Å
c	75.8 Å
Data	
Resolution	2.0 (2.11) Å
Nmeasured (unique)	46889 (14952)
Completeness (%)	99.3 (99.3)
Rmerge	6.3 (23.5) %
I/sigI	9.4 (3.1)

Model atoms	
Protein	1219
Water	200
Sulfate	15
Geometry	
Rmsd bond length	0.006 Å
Rmsd bond angles	1.51°
Rmsd dihedral angles	24.7°
Bav	24.6 Å²
Validation	
Rwork (free 5%)	19.5 (23.9) %
Est. error (cross-validated)	0.16 (0.18) Å



Resolution, # of measured waves = spots. Per atom, in x,y,z, at least 3 spots needed

Error values

THANK YOU!!